

Kurzbeitrag

Clusteranalyse als Instrument zur Gruppierung von spezialisierten Marktfruchtunternehmen

Michael Herink und Volker Petersen

Martin-Luther-Universität Halle-Wittenberg

Zusammenfassung

Ziel der Arbeit ist es, durch das Verfahren der Clusteranalyse landwirtschaftliche Unternehmen mit Marktfruchtbau mit ähnlichen Eigenschaften zu bündeln. Die Gruppenbildung soll unter anderem Hinweise darauf liefern, ob charakteristische Handlungsmuster mit der Höhe des Erfolges korrespondieren. Datengrundlage für die Untersuchung sind hoch spezialisierte Unternehmen mit Marktfruchtbau auf guten natürlichen Standorten in den Bundesländern Mecklenburg-Vorpommern und Brandenburg. Die hier durchgeführte hierarchisch-aggglomerative Clusteranalyse basiert im Wesentlichen auf dem Ward-Algorithmus. Die untersuchten Unternehmen werden in acht relativ homogene Gruppen unterteilt, die sich voneinander prägnant unterscheiden.

Folgende Ergebnisse sind festzuhalten. Eine Gruppe von Clustern mit einem insgesamt niedrigen Niveau der Naturalerträge setzt Aufwendungen ein, die erheblich, und zwar um 120 €/ha, voneinander abweichen. Das Gleiche gilt für eine zweite Gruppe von Clustern mit einem im Ganzen hohen Niveau der Naturalerträge. Umgekehrt erzielt eine nächste Gruppe von Clustern mit einem nahezu gleichen, niedrigen Aufwand Naturalerträge, die spürbar (11 dt/ha) auseinander fallen. Das Gleiche gilt für weitere Gruppe von Clustern mit einem überaus ähnlichen, aber sehr hohen Aufwand. Darüber hinaus zeigt sich, dass ein gegebener mittlerer bis hoher Erfolg auf betont verschiedenen Wegen zustande kommt. Von Gruppen, die mit sehr hohem Aufwand sehr hohe Naturalerträge erstellen, lassen sich Gruppen abgrenzen, die mit ausgeprägt niedrigem Aufwand mittlere bis hohe Naturalerträge realisieren. Ein Vergleich mit eindimensional angelegten Auswertungen auf der Grundlage von Erfolgsquartilen macht deutlich, dass diese Gruppierung Handlungen zur Steigerung des Erfolges impliziert, die im Lichte mehrdimensionaler Analysen nicht sachgerecht sind. Die Handlungslinien, um gute wirtschaftliche Ergebnisse zu erzielen, sind, anders als die eindimensionale Analyse nahe legt, recht divergent.

Was sind die Ursachen für die Beobachtungen? Die Analyse legt die Vermutung nahe, dass Unterschiede im Management und in den natürlichen Standortfaktoren die ungleichen Handlungsmuster und Erfolge begründen. Infolge der Datengrundlage muss jedoch offen bleiben, welche natürlichen Standortfaktoren eine hier nicht identifizierte Rolle gespielt haben könnten. Dem explorativen Ansatz der Analyse entsprechend lässt sich zudem nicht abgrenzen, durch welche Dimensionen des Managements die Ergebnisse ursächlich zu erklären sind. Die damit für diese und ähnlich gelagerte Unternehmen verbundenen Fragestellungen haben ein hohes Gewicht. Denn die Cluster mit den wirtschaftlich besten Ergebnissen weisen, teils divergent, auf Handlungsmuster hin, die auch unter Weltmarktbedingungen eine konkurrenzfähige Produktion ermöglichen.

Schlüsselwörter

Clusteranalyse; Unternehmensattribute; Management

1. Zielsetzung und Aufbau

In dem vorliegenden Beitrag soll eine heterogene Gruppe von landwirtschaftlichen Unternehmen mehrdimensional in homogene Fraktionen untergliedert werden. Zielsetzung ist

es herauszufinden, welche typischen Attribute diese Fraktionen kennzeichnen und ob sich daraus Hinweise auf Konstituenten des Erfolges herleiten lassen. Die mehrdimensionale Fraktionierung von Betrieben wird des Weiteren einer eindimensionalen Gliederung gegenüber gestellt. Durch den Vergleich sollen Aussagen darüber gewonnen werden, ob durch die mehrdimensionale Konstruktion von Einheiten Zusammenhänge, insbesondere im Hinblick auf die Konstituenten des Erfolges, zu identifizieren sind, die sich bei eindimensionaler Abgrenzung nicht erkennen lassen.

Um die mehrdimensionale Gruppenbildung sachgerecht durchzuführen, werden zunächst verschiedene clusteranalytische Verfahren mit ihren jeweils typischen Merkmalen vergleichend nebeneinander gestellt. Anschließend erfolgt die Auswahl eines für die vorliegende Fragestellung und den gegebenen Datenkranz geeigneten Algorithmus. Datengrundlage für die Clusteranalyse ist eine größere Anzahl von hoch spezialisierten Marktfruchtbetrieben. Die für die Untersuchung verwendeten zentralen Kenndaten dieser Unternehmen liegen vollständig für einen mehrjährigen Zeitraum vor. In einem weiteren Schritt werden die Betriebe dann in mehrere voneinander deutlich abgrenzbare Gruppen mit jeweils typischen Merkmalen unterteilt. Die charakteristischen Attribute der so gebündelten Unternehmen werden dabei vergleichend neben die Attribute jener Unternehmen gestellt, die sich bei eindimensionaler Schichtung auf der Grundlage von Erfolgsquartilen ergeben. Die Diskussion wendet sich insbesondere der Frage zu, ob die simultane Berücksichtigung mehrerer Kennzeichen zu zweckmäßigeren Aussagen über die Ursachen des Erfolges von Unternehmen führen.

2. Theoretischer Bezugsrahmen

Die vorliegende Analyse verzichtet darauf, im Vorwege Hypothesen über eventuelle Ursachen des Erfolges zu formulieren. Das Vorgehen ist also explorativ. Ein zur theoretisch nicht basierten Entdeckung von mehrdimensionalen Strukturen geeignetes Verfahren ist, sofern ratioskalierte Daten vorliegen, neben der explorativen Faktorenanalyse insbesondere die Clusteranalyse. Sie ist ein statistisches Verfahren, das Objekte einer ungeordneten Menge in repräsentative Gruppen einteilt. STEINHAUSEN und LANGER (1977) weisen darauf hin, dass dies nur dann möglich ist, wenn innerhalb der ungeordneten Objektmenge eine Struktur, eine Ordnung oder ein System von Ähnlichkeiten existiert.

2.1 Ein Literaturüberblick

In den 30er und 40er Jahren des zwanzigsten Jahrhunderts entwickelten HORST und GUTTMAN erste nutzbare Ansätze der unvollständigen Clusteranalyse (zitiert nach BACHER, 1996). Die unvollständige Clusteranalyse unterscheidet sich

von der heute weit verbreiteten deterministischen Clusteranalyse in dem Punkt, dass sie eine Zuordnung der Objekte im geometrischen Raum vornimmt, während die deterministischen Verfahren die Objekte mit Hilfe von mathematischen Algorithmen in definierte Gruppen einteilen. Die Geschichte der deterministischen Clusteranalyse ist eng an die technische Entwicklung leistungsstarker Rechner gekoppelt. Energische disziplinäre Impulse für die Entwicklung und Verbesserung der Clusteranalyse kamen aus der Soziologie, den Verhaltenswissenschaften und der Medizin. Hier besteht das Problem der Entdeckung von Zusammenhängen in teilweise sehr großen Objektmengen, deren Variablen oft multidimensional verknüpft sind. Eine erste Hinwendung auf den ländlichen und landwirtschaftlichen Bereich erfolgt durch KILCHENMANN und MOERGELI (1970). Sie untersuchen clusteranalytisch die wirtschaftsgeografische Struktur von 171 Schweizer Gemeinden mit Hilfe einer deterministischen Gruppenbildung unter Verwendung des so genannten Ward-Algorithmus. Eine der ersten agrarökonomisch basierten Clusteranalysen stellt SCHÄFFER (1972) vor. Er untersucht 96 Unternehmen, die er auf Grundlage von 17 ausgewählten Merkmalen in sieben bzw. acht charakteristische Gruppen einteilt. ALBACH et al. (1985) verwenden die Clusteranalyse, um kritische Wachstumsschwellen in der Unternehmensentwicklung zu identifizieren. Im Ergebnis charakterisieren sie das Verfahren als ein sehr verlässliches Instrument zur Abgrenzung von unterschiedlichen Unternehmenstypen. Im Mittelpunkt der Arbeit von GOTTSCHLICH (1995) steht die Schwachstellenanalyse von mehr als 2 000 größeren landwirtschaftlichen Unternehmen in Westfalen-Lippe. Im Zuge der Analyse stellt er vergleichend verschiedene Algorithmen der Clusteranalyse nebeneinander. Dabei stellt er fest, dass für große Objektmengen der Density-Linkage Algorithmus vergleichsweise gute Partitionen liefert. Die Objektmenge wird, unter Einsatz von sieben hauptsächlich Kennzahlen, in bis zu neun Cluster untergliedert. Diese Gruppen werden sodann vier Erfolgsklassen zugeordnet. Der Autor macht in diesem Rahmen auf die besonderen Datenprobleme aufmerksam, die mit der Nutzung von amtlichen Buchführungsergebnissen verbunden sind. Er fordert deshalb eine individuelle Überprüfung und gegebenenfalls Korrektur der dort ausgewiesenen Kennzahlen. REEP (1994) untersucht die Jahresabschlüsse landwirtschaftlicher Unternehmen der amtlichen Testbetriebsstatistik aus Sachsen auf Existenzgefährdung. Er erreicht, basierend auf dem K-means Algorithmus der Clusteranalyse, eine Einteilung in gefährdete und nicht gefährdete Unternehmen. DIECKMEIER (1996) nutzt die Clusteranalyse, um landwirtschaftliche Betriebe der Landkreise Emsland und Werra-Meißner anhand der Produktionsstruktur zu charakterisieren. Über den Zeitraum von 1970 bis 1990 gelingt es ihm, Wanderbewegungen zwischen Betriebsgruppen aufzuzeigen. Grundlage seiner Analyse ist der Ward-Algorithmus, der unter den dort gegebenen Voraussetzungen ein geeignetes Verfahren ist, um inhaltlich passende Einteilungen vorzunehmen. ADAMS (2004) untersucht mehr als 2 500 Unternehmen aus Rheinland-Pfalz auf ähnliche Eigenschaften. Der Gruppenbildung liegen insgesamt 13 Merkmale zugrunde. Zur Gruppenbildung werden vergleichend mehrere partitionierende Verfahren eingesetzt. Es gelingt ihr, auf Grundlage der aus Jahresabschlüssen gewonnenen Daten, und trotz der damit verbundenen Problematik (s.o.), insgesamt fünf relativ homogene Gruppen gegeneinander abzugrenzen.

Die Clusteranalyse bietet, wie alle weiteren multivariaten Verfahren auch, eine Vielzahl von Ansatzpunkten, um Einfluss auf die Ergebnisse zu nehmen. Stellvertretend genannt seien die Auswahl des Algorithmus, der verwendeten Kennzahlen oder der Anzahl von Clustern. Aussagen der Clusteranalyse sind demgemäß stets mit Blick auf diese Festlegungen zu sehen und zu bewerten.

2.2 Auswahl des Algorithmus

Grundsätzlich lassen sich die Clusterverfahren einteilen in graphentheoretische Verfahren, Optimierungsverfahren sowie in die weit verbreiteten partitionierenden und hierarchischen Verfahren (BACHER, 1996; BACKHAUS et al., 1996; EVERITT, 1993). Partitionierende Verfahren können wesentlich größere Objektmengen bearbeiten als die aufwändigeren Algorithmen der hierarchischen Verfahren, die zwischen allen Objekten aller Gruppen immer wieder neue Paarbeziehungen bestimmen. Da sie allerdings den großen Nachteil haben, keine Aussage über eine sinnvolle Gruppenzahl geben zu können, werden sie hier nicht eingesetzt. Die Gruppe der hierarchischen Verfahren ist gekennzeichnet durch die gute Nachvollziehbarkeit des Fusionierungsprozesses. Innerhalb der hierarchischen Verfahren werden vorwiegend die agglomerativen Algorithmen eingesetzt. Zu Beginn bildet jedes Objekt einen Cluster. In einem iterativen Prozess werden dann immer wieder Gruppen mit maximaler Ähnlichkeit zusammengeführt, bis die größte Partition erreicht ist.

In der vorliegenden Arbeit wird auf Grundlage der standardisierten Rohdaten eine Distanzmatrix erstellt. Distanzmaße bzw. Unähnlichkeitsmaße sind immer dann gegenüber Ähnlichkeitsmaßen vorzuziehen, wenn, wie hier, die Höhe und der absolute Abstand zwischen Objekten und nicht der Gleichlauf von Objektprofilen von Belang ist.

Zur Agglomeration von Objekten auf der Grundlage von Distanzmaßen stehen sieben eingeführte Algorithmen zur Verfügung. Hier wird das so genannte Ward-Verfahren ausgewählt. Grundlage dieses Verfahrens ist die Fehlerquadratsumme als Summe der internen Heterogenitätsmaße aller Cluster. Untersuchungen von BERGS (1981) zeigen, dass insbesondere dieses Verfahren im Vergleich mit den anderen Algorithmen sehr gute Partitionen liefert und die Objekte den „richtigen“ Gruppen zuordnet. Über besondere Vorteile verfügt das Verfahren unter anderem immer dann, wenn die Verwendung eines Distanzmaßes ein inhaltlich sinnvolles Kriterium ist, alle Variablen metrisch skaliert vorliegen, Ausreißer nicht vorliegen, die Variablen vertretbare Korrelationsmaße aufweisen, die Gruppen in etwa die gleiche Ausdehnung besitzen und die Objektanzahl in den Gruppen ungefähr gleich groß ist (BACKHAUS et al., 1996). Die entsprechenden Voraussetzungen sind in der vorliegenden Analyse durchweg gegeben. Empfindlich reagiert das Ward-Verfahren nicht nur gegenüber Ausreißern, sondern auch gegenüber kleinen Gruppen (vgl. MILLIGAN, 1967). Der Algorithmus neigt dann dazu, die kleinen Gruppen zusammenzufassen und so Gruppen ähnlicher Größe zu erzeugen. Bevor mit dem Ergebnis der Klassifikation weiter gearbeitet werden kann, ist somit eine inhaltliche Überprüfung der gebildeten Gruppen, eventuell durch den Vergleich mit der entsprechenden Gruppenbildung durch andere Algorithmen, unverzichtbar.

2.3 Festlegung der Gruppenanzahl

Das Endergebnis aller Clusteralgorithmen ist stets, dass alle Objekte in einem Cluster gebündelt sind. Da dies nicht das Ziel der Clusteranalyse sein kann, kommt der Frage, an welcher Stelle die Prozedur der Gruppenbildung unterbrochen werden muss, eine besondere Bedeutung zu. Techniken zur Festlegung einer sinnvollen Gruppenanzahl beruhen unter anderem auf dem Zuwachs der Fehlerquadratsumme als Maßstab der Heterogenität infolge sinkender Gruppenanzahl (Ward-Verfahren). Im Mittelpunkt dieser Analyse stehen, teils graphisch unterstützt, als eingeführte und gut nachvollziehbare Auswahlkriterien das Elbow-Kriterium und das Dendrogramm (BACKHAUS et al., 1996; BÜHL und ZÖFEL, 2002). Eine gute Gruppenanzahl wird unterschritten, wenn durch Wegnahme einer Gruppe der Zuwachs an Heterogenität erhebliche Ausmaße erreicht. Umgekehrt wird eine angemessene Gruppenanzahl überschritten, wenn die Heterogenität durch Hinzufügung einer weiteren Gruppe nicht mehr nennenswert reduziert wird und, in Verbindung damit, die Interpretierbarkeit der Ergebnisse nachhaltig erschwert wird.

2.4 Auswahl geeigneter Variablen

Für die Auswahl geeigneter Variablen sind zwei prinzipielle Überlegungen zu berücksichtigen. Die *erste* Überlegung betrifft die Bestimmung der Anzahl genutzter Variablen. Mit steigender Anzahl erhöht sich die Dimension der Untersuchung. Bei Analysen mit mehr als drei Variablen, dies entspricht der Verwendung von mehr als drei Dimensionen, wird eine Interpretation der Ergebnisse zunehmend schwierig.

Nach der Entscheidung für eine bestimmte Art und Anzahl von Variablen tritt unter Umständen das Problem der Korrelation auf. Damit ist, *zweitens*, über den tolerierbaren Grad an Korrelation zwischen einzelnen Größen zu entscheiden (Problem der Multikollinearität). Hohe Korrelationsgrade nehmen Einfluss auf die Ähnlichkeits- und Distanzmaße. Eine sehr enge Korrelation zwischen zwei Variablen könnte bedeuten, dass sich beide Variablen auf eine gemeinsame Hintergrundvariable zurückführen lassen. Im Prozess der Gruppenbildung wird dann eine Variable gegenüber den weiteren Variablen als wesentlich bedeutungsvoller gewichtet. Grenzwerte für noch tolerierbare Korrelationsgrade in hoch dimensionierten Analysen liegen zwischen 0,4 und 0,5 (NIEHAUS, 1987; PYTLIK, 1995). Teilweise werden auch deutlich höhere Toleranzwerte genannt (BACKHAUS, 1996). Insbesondere bei leicht überschaubaren niedrig dimensionierten Analysen lassen sich auch höhere Grenzwerte vertreten; sie führen dort in der Regel nicht zu einer Verzerrung und Fehlinterpretation der Ergebnisse (SPILKE, 2004).

3. Datengrundlage

Empirische Grundlage der Analyse sind hoch spezialisierte Unternehmen mit Marktfruchtbau in Mecklenburg-Vorpommern und Brandenburg. Eine begrenzte Diversifikation findet durch die Bereitstellung von Dienstleistungen in enger Anbindung an das Kerngeschäft statt. Die Unternehmen zeichnen sich durch über dem Durchschnitt liegende Produktionsverhältnisse und vergleichsweise günstige natürliche Standortverhältnisse aus. Infolge der Mitglied-

schaft in einem überregional organisierten Beratungsring mit hoher Beratungsintensität sind auch die Managementverhältnisse als überdurchschnittlich einzustufen. In die Analyse fließen 24 Unternehmen ein. Die für diese Betriebe vorliegende langjährige Datenreihe wurde mit Blick auf die hier durchgeführte Untersuchung für die Wirtschaftsjahre 2000/01 bis 2002/03 ausgewertet. Um einzeljährige Besonderheiten zu glätten, wurden alle dreijährigen Daten zu Durchschnittswerten zusammengefasst. Die für die nachfolgende Gruppenbildung im Vorwege festgelegten Kennzahlen sind der gesamte Aufwand (ohne Ansatz für die Bindung von Anlage- und Umlaufvermögen), der Reinertrag und der Naturalertrag, abgebildet durch die Getreideeinheit als Spitzenkennzahl.¹ Grundlage für die Kennzahlen sind die jeweiligen Jahresabschlüsse der Unternehmen. Deren Kennzahlen sind durch eine entsprechende individuelle Aufbereitung in Zusammenarbeit mit der Beratung und den Unternehmen um Angaben korrigiert, die im Hinblick auf andere (z.B. steuerliche) Zielsetzungen erfolgt sind oder auch aus eventuellen Fehlbuchungen entstanden sind. Beispielhaft seien hier die Aktivierung von Aufwendungen für die Unterhaltung, die mittelfristige Verteilung und die sachliche Zuordnung von außerordentlichen Erträgen sowie die Bereinigung von Kennziffern des Marktfruchtbaus um eventuelle, nicht zutreffend zugeordnete Bestandteile genannt. Ein Verzicht auf diese Korrekturen kann empirische Analysen nachdrücklich erschweren. Die so aufbereiteten Kennzahlen liegen lückenlos für den gesamten Untersuchungszeitraum vor.

4. Empirische Analyse

4.1 Modelleigenschaften

Ein im Rahmen der Clusteranalyse verwendetes wesentliches Kriterium zur Überprüfung der statistischen Modellqualität (Reliabilität) ist der F-Wert. Er gibt an, wie hoch die Streuung einer Variablen in einem Cluster im Verhältnis zu ihrer Streuung in der gesamten Datenmenge ist. Der F-Wert ist deshalb ein Indikator für die Homogenität der gefundenen Cluster. Die hier empirisch gemessenen F-Werte überschreiten in keinem Fall den Wert von 0,31. Sie liegen damit stets weit unterhalb des kritischen F-Grenzwertes von 1,0. Das Modell fasst die Unternehmen also gut in homogenen Gruppen zusammen. Zu den Korrelationskoeffizienten: Sie bewegen sich hier zwischen unteren und oberen Eckwerten von 0,066 und 0,699. Trotz der recht engen Korrelation von 0,699 zwischen Reinertrag und Naturalertrag ist vor dem Hintergrund der niedrig dimensionierten Analyse gleichwohl nicht von einer Verletzung der Modellprämissen und einer Verzerrung der Ergebnisse auszugehen (s.o.).

Der hier verwendete Ward-Algorithmus neigt im Gegensatz zu anderen Algorithmen dazu, kleine, relativ gut abgrenzbare Cluster selbst dann zu fusionieren, wenn eine Vernetzung sachlich nicht angemessen erscheint. In dem vorliegenden Fall signalisiert das Elbow-Kriterium, dass grundsätzlich eine gut nutzbare Clusteranzahl zwischen fünf und acht Clustern liegt. Dabei zeichnet sich eine Lösung mit

¹ Eine vergleichende Darstellung verschiedener Methoden zur Abbildung der natürlichen Gegebenheiten in der Getreideeinheit als Spitzenkennzahl findet sich bei HERINK (2004).

sechs Clustern als günstig ab, da weitere Gruppenbildungen nicht mehr nennenswert zu einer Abnahme der kumulierten Heterogenität beitragen. Nun wird hier allerdings ein mit dem Ward-Algorithmus verbundenes Problem offenkundig: Das Ward-Modell verknüpft auf der Grundlage von sechs Clustern Objekte miteinander, die von allen anderen Algorithmen eindeutig getrennt werden.² Es führt erst bei einer Lösung mit acht Clustern zu einer auch sachlich angemessenen Trennung von Betrieben. Trotz der damit verbundenen Problematik einer größeren Unübersichtlichkeit und aufwändigeren Interpretierbarkeit soll deshalb ein Modell mit acht Clustern erstellt werden.

4.2 Ergebnisse der Clusteranalyse

4.2.1 Merkmale der Gruppen

Zentrale Aussagen der Analyse sind in Tabelle 1 zusammengestellt. Das eingangs genannte vergleichsweise hohe produktionstechnische und wirtschaftliche Niveau der ausgewerteten Betriebsgruppe wird durch die im Durchschnitt angeführten Kenndaten quantitativ untermauert. Zu den Clustern: Die hervorstechenden Eigenschaften jedes Clusters sind symbolhaft ausgewiesen. Die erfassten Unternehmen werden jeweils einem der acht Cluster zugeordnet. Die Größe der entschlüsselten Cluster ist allerdings recht heterogen. Die Anzahl der zugeordneten Objekte schwankt zwischen einem und sieben Unternehmen.

Tabelle 1. Ergebnisse und Rangierung im 8-Cluster-Modell

	Ergebnis ^a			Merkmale		
	Reinertrag	Naturalertrag	Aufwand	Reinertrag	Naturalertrag	Aufwand
Einheit	€/ha	dt/ha	€/ha	€/ha	dt/ha	€/ha
Cluster 1	936	101,5	763	+++	+++	---
Cluster 2	627	90,3	840	++	+	--
Cluster 3	410	78,9	867	--	--	-
Cluster 4	422	95,8	1191	-	++	+++
Cluster 5	525	90,5	937	Ø	+	Ø
Cluster 6	445	85,0	960	-	-Ø	Ø
Cluster 7	668	98,1	1087	++	++	++
Cluster 8	377	85,2	1084	---	-Ø	++
Durchschnitt	496	87,0	948			

^aDie Eckwerte jedes Merkmals sind unterstrichen.

Quelle: Eigene Berechnungen

Das Cluster 1, es besteht aus nur einem Unternehmen, zeichnet sich durch den höchsten Reinertrag, den höchsten Naturalertrag und den geringsten Aufwand aus.³ Der Reinertrag von 936 €/ha liegt um 88 v.H. und der Naturalertrag mit 101,3 dt/ha um 16 v.H. über dem Durchschnitt der gesamten Unternehmensgruppe. Zugleich unterschreitet der Aufwand mit 763 €/ha um rund 20 v.H. das Gruppenmittel. Dieser bemerkenswerten Leistung entsprechend bildet das Unternehmen auch in den getesteten Centroid-, Median-, Single-Linkage- und Average-Linkage-Verfahren im Falle einer Lösung mit nur zwei Clustern stets einen eigenen Cluster. Das Cluster 2 ist ebenfalls gekennzeichnet durch

einen hohen Reinertrag von 627 €/ha, einen leicht überdurchschnittlichen Naturalertrag von 90,3 dt/ha, allerdings verbunden mit einem unterdurchschnittlichen Aufwand von 840 €/ha. Das Cluster 3 weist desgleichen einen unterdurchschnittlichen Aufwand von 867 €/ha auf. Da zugleich jedoch mit 78,9 dt/ha das niedrigste Ernteergebnis realisiert wird, grenzt sich dieses Cluster mit einem Reinertrag von nur 410 €/ha deutlich stark vom vorigen Cluster 2 ab. Das Cluster 4 realisiert ebenfalls einen geringen Reinertrag von nur 422 €/ha. Hier besteht also kaum ein Unterschied zum Cluster 3. Deutliche Unterschiede sind jedoch im Hinblick auf das Zustandekommen dieses Ergebnisses zu erkennen. Während Cluster 3 ein sehr geringer Naturalertrag charakterisiert, wird hier mit 95,8 dt/ha ein bemerkenswert guter Naturalertrag erzielt. Kennzeichen dieses Clusters, und das führt zu der eher niedrigen Rentabilität, ist der außerordentlich hohe Aufwand von 1 191 €/ha. Die Cluster 5 und 6 bewegen sich mit ihren Kennzahlen Reinertrag, Naturalertrag und Aufwand gerade auf dem durchschnittlichen Niveau aller Unternehmen. Jedoch ist festzustellen, dass das Cluster 5 im Unterschied zum Cluster 6 in allen Kennzahlen geringfügig günstigere Werte aufweist. Im Ergebnis liegt so der Reinertrag des Clusters 5 mit 525 €/ha um 29 €/ha über dem mittleren Reinertrag von 496 €/ha und der Cluster 6 mit 445 €/ha um 51 €/ha darunter. Der Cluster 7 weist gewisse Ähnlichkeiten mit dem Cluster 4 auf, bedingt

durch die in beiden Fällen sehr hohen Naturalerträge und Aufwendungen. Allerdings, und das sind die Unterschiede, übertrifft hier der Ertrag mit 98,1 dt/ha (zweitbestes Ernteergebnis) und unterschreitet der Aufwand mit 1 087 €/ha (zweithöchster Aufwand) die korrespondierenden Positionen in Cluster 4. Demgemäß ergeben sich zwischen beiden Clustern Unterschiede im Reinertrag von 246 €/ha. Das siebente Cluster realisiert mit diesem Gefüge die zweithöchsten Reinerträge. Den niedrigsten Reinertrag mit nur 377 €/ha erwirtschaftet das Cluster 8. Hier begegnen sich ein mittlerer bis leicht unterdurchschnittlicher Naturalertrag von

85,2 dt/ha und ein hoher Aufwand mit 1 084 €/ha.

Zusammenfassend ist festzuhalten, dass die gebildeten Gruppen in sich geschlossene Merkmale aufweisen und sich getrennten Gruppen auch sachlich gut zuordnen lassen.

4.2.2 Gruppenmerkmale im Vergleich

Die typischen Merkmale der konstruierten Gruppen geben erste Hinweise auf mögliche Unterschiede in den Eigenschaften des natürlichen Standortes und des Managements. Dies soll an drei Beispielen illustriert werden.

Cluster 2 und Cluster 4: Die Unternehmen des Clusters 2 bewegen sich mit einem Naturalertrag von 90,3 dt/ha gegenüber dem vierten Clusters mit 95,8 dt/ha auf einem deutlich geringeren Ertragsniveau. Kennzeichnend für das vierte Cluster ist daneben der mit 1 191 €/ha höchste Aufwand. Dabei zeigt die Tiefenanalyse des Datenmaterials,

² Als Referenzmodelle wurden hier das Centroid-, das Median-, das Single-Linkage- sowie die beiden Average-Linkage-Verfahren eingesetzt.

³ Die Kennzahlen beziehen sich stets auf einen Hektar.

dass sich dieser hohe Aufwand in allen Bereichen widerspiegelt, insbesondere bei den Abschreibungen für den groß dimensionierten Maschinenpark. Eine dafür mögliche Ursache sind die hier angebotenen Dienstleistungen. Allerdings liegt der Aufwand auch nach Berücksichtigung eines Korrekturfaktors immer noch deutlich über den Aufwendungen der Vergleichsgruppe.⁴ Alles in allem werden mit dem ertragsextensiveren zweiten Clusters um 205 €/ha höhere Reinerträge realisiert. Die dortige Formation mit leicht überdurchschnittlichen Naturalerträgen und ausgeprägt niedrigen Aufwendungen führt im Vergleich aller Cluster zu der drittbesten Rangierung. Cluster 6 und 8: Die Naturalerträge beider Cluster liegen auf leicht unterdurchschnittlichem Niveau. Sie weichen des Weiteren mit 85,0 dt/ha (Cluster 6) und 85,2 dt/ha (Cluster 8) kaum voneinander ab. Gemessen an diesem Indikator sind also beide Gruppen gut vergleichbar. Allerdings weist das Cluster 8 einen um 124 €/ha höheren Aufwand auf. Entsprechend niedriger ist der Reinertrag. Cluster 2 und 5: Die Naturalerträge beider Cluster sind im höheren Ertragsbereich angesiedelt; sie sind zudem mit 90,3 (Cluster 2) und 90,5 dt/ha (Cluster 5) weitgehend deckungsgleich. Dennoch unterscheiden sich die Aufwendungen um 97 €/ha zu Ungunsten des fünften Clusters. Parallel dazu ist der Reinertrag des zweiten Clusters um ca. 100 €/ha höher.

Die vorgestellte Analyseverfahren ist explorativ angelegt. Sie kann insofern auch nur erste Hinweise auf eventuelle Gründe für Unterschiede im Erfolg geben und somit Ausgangspunkt für die weiter gehende Erforschung von Ursachen in der untersuchten Gruppe von Unternehmen oder in ähnlichen Unternehmen sein.

In einem künftigen Ansatz sollte das hauptsächliche Augenmerk auf zwei Aspekte gerichtet sein. Wichtige Einflussfaktoren auf die Höhe der Erträge und die dazu notwendigen Aufwendungen sind *erstens* die Produktivität des Bodens, regionale klimatische Ertragsfaktoren oder die Bearbeitbarkeit der Böden. Der zur Abbildung der natürlichen Standorteigenschaften im Jahresabschluss ausgewiesene Vergleichswert ist dabei ein denkbar ungeeigneter Indikator. Obwohl sich die Vergleichswerte der Unternehmen in einer weiten Bandbreite bewegen, lässt sich keine ursächliche Beziehung zu den Ertrags- und Aufwandsgrößen nachweisen. Um tiefer gehende Informationen zu gewinnen, sind also differenziertere Erhebungen und Einstufungen notwendig. Ein gleichfalls bedeutsamer Einflussfaktor sind *zweitens* die Zielvorstellungen, die Ausrichtung und die Qualität der Unternehmensführung (PETERSEN, 2003). Hierzu liegen weiter gehende Informationen für die untersuchten Unternehmen bislang nicht vor. Da sie sich

durch eine Auswertung der Ergebnisse von Jahresabschlussdaten nur begrenzt gewinnen lassen, sind gleichfalls zusätzliche Erhebungen angebracht.

4.2.3 Merkmale eindimensionaler Gruppenbildung

Grundlage für die eindimensionale Gruppenbildung ist, entsprechend der Verfahrensbezeichnung, nur eine Variable. Der Tabelle 2 liegt eine einfache Gruppenbildung gemäß der unterschiedlichen Erfolgsquartile zugrunde. Gemessen am Reinertrag wird exemplarisch das Viertel der erfolgreichen Unternehmen mit dem der abfallenden Unternehmen verglichen.

Tabelle 2. Eindimensionale Erfolgsanalyse anhand der Quartileinteilung

	Unternehmen	Reinertrag in €/ha	Naturalertrag in dt/ha	Aufwand in €/ha
Oberes Quartil	1	936	101,5	763
	11	668	98,1	1087
	7	636	88,8	881
	3	634	92,8	840
	2	613	89,4	799
	6	561	90,3	949
Durchschnitt oberes Quartil		675	93,5	887
Unteres Quartil	20	415	79,5	863
	13	396	82,6	957
	21	369	80,5	1091
	18	350	73,8	907
	24	344	88,8	1098
	23	330	82,7	994
Durchschnitt unteres Quartil		368	81,3	985
Durchschnitt aller Quartile		496	87,0	948

Quelle: Eigene Berechnungen

Die so vorgenommene Gruppenbildung birgt die Gefahr von Fehlansagen in sich. Beispielfhaft seien zwei Richtungen genannt. (1) So wird das Unternehmen 11 in das Quartil der erfolgreichsten Unternehmen eingeordnet. Nun übertrifft sein Aufwand den Durchschnitt dieser Gruppe erheblich, dadurch wird zunächst der mittlere Aufwand der Gruppe deutlich angehoben. Dennoch könnte man zu dem falschen Ergebnis gelangen, dass erfolgreiche Unternehmen stets mit einer niedrigeren Organisationsintensität (887 zu 985 €/ha) wirtschaften. Eine Aussage, die durch die Ergebnisse der Clusteranalyse (Cluster 7) nicht bestätigt wird. (2) Eine zweite Fehldeutung ergibt sich aus dem Vergleich der Naturalerträge im oberen Quartil (93,5 dt/ha) mit denjenigen im unteren Quartil (81,3 dt/ha). Der Vergleich impliziert, dass erfolgreiche Unternehmen sich stets durch besonders hohe Naturalerträge auszeichnen, also eine Zunahme der Naturalerträge auf eine Steigerung des Erfolges hinwirkt. Die Clusteranalyse weist auch in dieser Hinsicht auf eine überaus vielschichtige Lage hin (vgl. Cluster 2 und 4 sowie Cluster 3 und 8). Die mehrdimensionalen Strukturen der Clusteranalyse sind in dem vorliegenden Fall gut geeignet, eher einseitigen Deutungen entgegen zu treten.

5. Einordnung

Die Gruppenbildung soll unter anderem Hinweise darauf liefern, welche Attribute Unternehmen mit gleichem bzw.

⁴ Diese Einnahmen aus den Dienstleistungen betragen 132 €/ha. Wird unterstellt, dass diese Summe den nicht separat aufgeführten Aufwendungen für die Dienstleistungen entspricht, so liegt der Aufwand immer noch um 219 €/ha über den Aufwendungen des zweiten Clusters.

unterschiedlichem Erfolg kennzeichnen. Unter Berücksichtigung der eingeführten Kriterien werden die untersuchten Unternehmen in acht relativ homogene Gruppen unterteilt, die sich voneinander prägnant unterscheiden. Folgende Ergebnisse sind festzuhalten.

Eine Gruppe von Clustern mit gleichem, *niedrigem Niveau der Naturalerträge* arbeitet mit Aufwendungen, die sich erheblich (120 €/ha) voneinander unterscheiden. Das Gleiche gilt für eine zweite Gruppe von Clustern mit *gleichem*, aber *hohem Niveau der Naturalerträge*.

Eine nächste Gruppe von Clustern erzielt mit nahezu gleichem, *niedrigem Aufwand* Naturalerträge, die spürbar (11 dt/ha) auseinander fallen. Das Gleiche gilt für eine Gruppe von Clustern mit nahezu gleichem, aber sehr *hohem Aufwand*.

Darüber hinaus zeigt sich, dass ein *gegebener* mittlerer bis hoher *Erfolg* auf verschiedenen Wegen zustande kommt. Von Gruppen, die mit sehr hohem Aufwand sehr hohe Naturalerträge erstellen, lassen sich Gruppen abgrenzen, die mit ausgeprägt niedrigem Aufwand mittlere bis hohe Naturalerträge realisieren. Das Gleiche gilt sinngemäß für Gruppen mit einem gegebenen, geringen Erfolg. Ein Vergleich mit eindimensional angelegten Auswertungen auf der Grundlage von Erfolgsquartilen macht deutlich, dass diese Gruppierung Handlungen zur Steigerung des Erfolges impliziert, die im Lichte mehrdimensionaler Analysen nicht sachgerecht sind.

Was sind die Ursachen für die Beobachtungen? Die Analyse legt die Vermutung nahe, dass Unterschiede im Management und in den natürlichen Standortfaktoren die ungleichen Handlungsmuster und Erfolge begründen. Infolge der Datengrundlage muss jedoch offen bleiben, welche natürlichen Standortfaktoren eine hier nicht identifizierte Rolle gespielt haben könnten. Dem explorativen Ansatz der Analyse entsprechend lässt sich zudem nicht abgrenzen, durch welche Dimensionen des Managements die Ergebnisse ursächlich zu erklären sind. Die damit verbundenen Fragen sind für diese für diese und ähnlich gelagerte Unternehmen noch zu klären.

Literatur

- ADAMS, A. (2004): Erfolgsanalyse von landwirtschaftlichen Unternehmen im südlichen Rheinland-Pfalz. Halle (Saale) (unveröffentlicht).
- ALBACH, H., K. BOCK und T. WARNKE (1985): Kritische Wachstumsschwellen in der Unternehmensentwicklung. In: Schriften zum Mittelstand. Neue Folge, Heft 7, Bonn.
- BACHER, J. (1996): Clusteranalyse: Anwendungsorientierte Einführung. Oldenbourg Verlag, München.
- BACKHAUS, K. et al. (1996): Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Lehrbuch, 8. Auflage. Springer Verlag Berlin, Heidelberg, New-York.
- BERGS, S. (1981): Optimalität bei Cluster-Analysen. Dissertation, Fachbereich Wirtschafts- und Sozialwissenschaften, Westfälische Wilhelm-Universität Münster.
- BÜHL, A. und P. ZÖFEL (2002): SPSS 11. Einführung in die moderne Datenanalyse unter Windows. Pearson Education Deutschland, München.

- DIEKMEIER, R. (1996): Entwicklung der landwirtschaftlichen Produktionsstrukturen in den Kreisen Emsland und Werra-Meißner unter besonderer Berücksichtigung der EU-Agrarreform. In: Interdisziplinäre Studien zur Entwicklung in ländlichen Räumen. Band 12, Wissenschaftsverlag Vauk, Kiel.
- EVERITT, B.-S. (1993): Cluster Analysis. Third Edition. Halsted Press, New York & Edward Arnold, London.
- GOTTSCHLICH, R. (1995): Empirische Identifikation von typischen Schwachstellenprofilen Landwirtschaftlicher Unternehmen: Analysemöglichkeiten auf der Grundlage von Buchführungsdaten. Forschungsgesellschaft für Agrarpolitik und Agrarsoziologie e.V., Bonn.
- GUTTMANN, L. (1941): The Quantification of a Class of Attributes: A Theory and Method of Scale Construction. In: Horst, P. et al.: The Prediction of Personal Adjustment. New York: 319-348.
- HERINK, M. (2004): Operatives Management als Erfolgsfaktor in spezialisierten Marktfruchtbetrieben – eine empirische Analyse. Diplomarbeit, Landwirtschaftliche Fakultät, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale).
- HORST, P. (1935): Measuring Complex Attitudes. In: Journal of Social Psychology 6: 369-374.
- KILCHENMANN, A. und W. MOERGELI (1970): Typisierung der Gemeinden im Kanton Zürich mit multivariaten statistischen Methoden auf Grund ihrer wirtschaftsgeographischen Struktur. Naturforschende Gesellschaft in Zürich.
- MILLIGAN, G.-W. (1967): Some Methods for Classification and Analysis of Multivariate Observation. Proceeding of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley.
- NIEHAUS, H.-J. (1987): Früherkennung von Unternehmenskrisen – Die statistische Jahresabschlussanalyse als Instrument der Abschlussprüfung. Institut für Revisionswesen der Westfälischen Wilhelm-Universität Münster, hrsg. von Jörg Baetge, Düsseldorf.
- PETERSEN, V. (2003): Unternehmensführung und Unternehmenserfolg. In: Agrarwirtschaft 52 (2): 107-117.
- REEP, M. (1994): Untersuchungen zur Existenzgefährdung wieder-eingerichteter Landwirtschaftsbetriebe anhand statistischer Analysemethoden. Diplomarbeit, Landwirtschaftliche Fakultät, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale).
- PYTLIK, M. (1995): Diskriminanzanalyse und künstliche neuronale Netze zur Klassifizierung von Jahresabschlüssen – ein empirischer Vergleich. In: Europäische Hochschulschriften, Reihe 5, Volks- und Betriebswirtschaft, Band 1688. Verlag P. Lange, Frankfurt (Main).
- SCHÄFFER, K.-A. (1972): Klassifizierung landwirtschaftlicher Betriebe mit Hilfe multivarianter statistischer Verfahren. In: Agrarstatistische Studien 1972, Heft 10, Statistisches Amt der Europäischen Gemeinschaften: 5ff.
- SPIELKE, J. (2004): Persönliche Mitteilungen. Halle (Saale).
- STEINHAUSEN, D. und K. LANGER (1977): Clusteranalyse: Einführung in Methoden und Verfahren der automatischen Klassifikation. De Gruyter, Berlin.

Kontaktautor:

PROF. DR. VOLKER PETERSEN
Martin-Luther-Universität Halle-Wittenberg, Landwirtschaftliche Fakultät,
Institut für Agrarökonomie und Agrarraumgestaltung
06099 Halle (Saale)
Tel.: 03 45-55 22 362, Fax: 03 45-55 27 110
E-Mail: petersen@landw.uni-halle.de